DeepProZyme User Guide

Overview

DeepProZyme is a web-based platform for protein function prediction. It supports sequence-level deep learning tools, including:

DeepTFactor: Predicts whether a protein sequence encodes transcription factor (TF), supporting regulatory network reconstruction.

DeepECtransformer: Annotates enzyme commission (EC) numbers from protein sequences, allowing the functional annotation of uncharacterized genes.

This page explains:

- Section 1: How to run analyses with each DeepProZyme tools
- Section 2: How to interpret the results

Section 1: How to run analyses with DeepProZyme tools

Step 1. Enter your job name.

Step 2. (Optional) Enter your email address.

You may provide an email address to receive updates when your analysis:
begins,
finishes successfully, or
encounters an error.

Step 3. Provide your protein sequence (FASTA format).

You may paste the sequence directly into the input box, or upload a FASTA file. A sample FASTA file is available for reference to ensure correct formatting. For DeepECtransformer, you may additionally request sequence logo generation when a single protein sequence is entered. Moreover, the amino acid length that can be analysed by DeepECtransformer is limited to 2,500 amino acids, whereas DeepTFactor supports the analysis of protein sequences up to 1,000 amino acids in length.

Step 4. Submit your analysis.

Once all required fields are completed, click Submit. Your analysis will be placed in the processing queue.

Section 2: How to interpret the results

2.1. Interpreting results from DeepTFactor

Output file: DeepTFactor_1.csv

1) DeepTFactor_1.csv

Column	Meaning

Sequence ID	Identifier for the input protein sequence
Prediction	Predicted as Transcription Factor (TF) or non-TF
Score	Prediction score

2.2. Interpreting results from DeepECtransformer

Output file: DeepECtransformer.csv

1) DeepECtransformer.csv

Column	Meaning
Sequence ID	Identifier for the input protein
Prediction	Predicts EC number
Score	Prediction score. If DeepECtransformer cannot assign an EC number, the sequence is listed as unassigned, as no DIAMOND-based annotation is added in such cases.